

[PDF] High Performance Spark: Best Practices For Scaling And Optimizing Apache Spark

Holden Karau, Rachel Warren - pdf download free book



Books Details:

Title: High Performance Spark: Best

Author: Holden Karau, Rachel Warren

Released:

Language:

Pages: 358

ISBN: 1491943203

ISBN13: 9781491943205

ASIN: 1491943203

[CLICK HERE FOR DOWNLOAD](#)

pdf, mobi, epub, azw, kindle

Description:

Apache Spark is amazing when everything clicks. But if you haven't seen the performance improvements you expected, or still don't feel confident enough to use Spark in production, this

practical book is for you. Authors Holden Karau and Rachel Warren demonstrate performance optimizations to help your Spark queries run faster and handle larger data sizes, while using fewer resources.

Ideal for software engineers, data engineers, developers, and system administrators working with large-scale data applications, this book describes techniques that can reduce data infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you'll also learn how to make it sing.

With this book, you'll explore:

- How Spark SQL's new interfaces improve performance over SQL's RDD data structure
 - The choice between data joins in Core Spark and Spark SQL
 - Techniques for getting the most out of standard RDD transformations
 - How to work around performance issues in Spark's key/value pair paradigm
 - Writing high-performance Spark code without Scala or the JVM
 - How to test for functionality and performance when applying suggested improvements
 - Using Spark MLlib and Spark ML machine learning libraries
 - Spark's Streaming components and external community packages
-

- Title: High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark
 - Author: Holden Karau, Rachel Warren
 - Released:
 - Language:
 - Pages: 358
 - ISBN: 1491943203
 - ISBN13: 9781491943205
 - ASIN: 1491943203
-